Heterogeneous & Homogeneous & Bio-
CHEMCATCHEM
CATALYSIS

Supporting Information

Engineered Bacterial Mimics of Human Drug Metabolizing Enzyme CYP2C9

Andrea Rentmeister, Tristan R. Brown, Christopher D. Snow, Martina N. Carbone, and Frances H. Arnold*[a]

cctc_201000452_sm_miscellaneous_information.pdf

1 Materials and Methods

1.1 Cloning

Restriction enzymes were purchased from NEB. High fidelity polymerase was Phusion (Finnzymes). DNA purification kits were from Qiagen and Zymo. Taq polymerase for EP-PCR was from Applied Biosystems. GeneMorphII Random Mutagenesis kit was from Stratagene. Primers were ordered from Integrated DNA Technologies at desalted grade. Sequencing was done by Laragen.

1.1.1 Reattaching the heme and reductase domains

The heme domain of 13C9 was amplified by PCR1: Template pcWORI-peroxygenase-13C9, 5'-primer: BamH1fwd: 5'-CACAGGAAACAGGATCCATCGTGCTTAGG-3'; 3'-primer: 13Sacrev: 5'-GCTTTTATCACAAAGCCTTTAGGTTTTAACGTTAGAGTTTCTTCAATATCGAGCTCGTAG -3'. The reductase domain of $P450_{BM3}$ was amplified by PCR2: Template pcWORI-P450BM3, 5'-primer: 13Sacfwd 5'-CTACGAGCTCGATATTGAAGAAACTCTAACGTTAAAACCTAAAGGCTTTGTGATAAAAGC-3'; 3'-primer: red+44: 5'-GGCTCAGATCTGCTCATGTTTGACAGCTTATC-3'. Assembly PCR: Template PCR1 and PCR2, 5'primer: BamH1fwd, 3'-primer: red+44. The product was digested using 1 uL Dpn I for 2 h and purified from an agarose gel using a gel extraction kit (QIAGEN). The product was then cloned into pcWORI using BamHI and EcoRI restriction sites. The resulting construct was called 13C9R1 and transformed into electrocompetent *E. coli* DH5α.

1.1.2 Cloning of pET constructs for crystallization

Heme domains of selected monooxygenase variants 13C9R1, 22A3, and X3H1 were cloned into pET22b+ providing a C-terminal hexahistidine tag. To do so, an internal NdeI site had to be removed and an XhoI site had to be introduced, which was done by overlap extension PCR. PCR1: Template pcWORI-monooxygenase-variant (e.g. X3H1), 5'-primer HEMEfwd1: 5'-CAGGAAACAGGATCAGCTTACTCCC-3'; 3'-primer HEMENDE1MUTREV: 5'-CAGGAGCAGTTGGCCAgATGCGCAGCGCTT-3'. PCR2: Template pcWORI-monooxygenase-variant (e.g. X3H1), 5'-primer HEMENDE1MUTFWD: AAGCGCTGCGCATCTGGCCAACTGCTCCTG; 3'-primer HEMEREVXhoI: 5'-ACGACTCGAGAGTGCTAGGTGAAGGAATAC-3'. Assembly PCR: PCR1 and PCR2 with primers HEMEfwd1 and hemerevXhoI. Resulting products were digested with NdeI for 2 hours, purified and cloned into pET22b+ (Invitrogen) using NdeI and XhoI-sites. The resulting pET22b+ constructs were transformed into electrocompetent *E. coli* BL21.

1.1.3 Generation of l ibraries

Error-prone PCR was used to create random mutant libraries. Primers amplify the heme domain: 5'-primer: hemefwd1: 5'-CAGGAAACAGGATCAGCTTACTCCC-3'; 3'-primer: hemerev1: 5'-CGCGCCGTTCCTTCAGCTGTTCCC-3'. In the first two rounds the mutagenesis Gene Morph II kit (Strategene) was used according to the manufacturer's instructions. For rounds 3-5, EP-PCR with Taq-polymerase and $MnCl_2$ was used. Typically 4 error-prone PCRs with $MnCl_2$ concentrations between 50 µM and 500 µM were performed. Amplified DNA was Dpn I digested, gel purified and digested by BamHI and SacI, followed by gel purification. DNA was recovered using a gel extraction kit (Qiagen) and ligated into the vector pcWORI that had been digested and purified in the same way. A 96-well test plate

of the resulting libraries was expressed and analyzed by CO-binding. To determine the error-rate, 5-10 clones were sequenced.

1.2 Screening

1.2.1 High-throughput screening with purpald assay

Expression of $P450_{BM3}$ variants in 96-well plates and preparation of cell lysates were performed as described.[1] Screening of enzyme activity on naproxen **(1)** was carried out in 96-well microtiter plates mixing 90 µL KPi (100 mM, pH 8.0), 50 µL cell lysate, 10 µL 100 mM of **1** in DMSO, and 50 µL KPi 2 mM NADPH. Reaction mixtures were incubated for 60 minutes at room temperature and then combined with 50 µL 60 mM purpald (Sigma) in 2 M NaOH. Absorbance at 550 nm was recorded after 30–120 minutes. The most active variants identified in the colorimetric screen were re-tested in quadruplicate.

Winners were defined by improved activity in the screen compared to parent, which was cultured and tested on the same plate. Thresholds were chosen between 1.3 and 1.7-fold improvement. Winners were rescreened in quadruplicate. The most active variants were expressed and purified for bioconversions and subsequent HPLC analysis. The parent for the next round was chosen according to the turnover in bioconversions.

1.2.2 High-throughput screening with AAP assay

Screening of enzyme activity on 1-naphthalene acetic acid (**8**) was carried out in 96-well microtiter plates mixing 40 µL KPi (100 mM, pH 8.0), 100 µL cell lysate 10 µL 100 mM of **8** in DMSO, and 50 µL KPi 2 mM NADPH. Reaction mixtures were incubated for 60 minutes at room temperature and then combined with 60 µL urea (8M in 200 mM NaOH) and 18 µL 4-amino antipyrine (4-AAP) (1.2 g/100 mL), and 18 µL potassium persulfate ($K_2S_2O_8$) (1.2 g/100 mL). Absorbance at 510 nm was recorded after 5–30 minutes.

1.2.3 Small-scale reactions

Small-scale reactions were carried out in 600 µL scale using 1 mM substrate (naproxen or ibuprofen), 0.5 µM P450, 400 µM NADPH (Codexis), 20 mM glucose-6-phosphate (Sigma) and 1 U/mL glucose-6-phosphate dehydrogenase (USB). After stirring for 16 hours at room temperature, the reaction mixtures were combined with 600 µL acetonitrile, centrifuged at 14,000 rpm for 10 min, filtered, and analyzed by HPLC or HPLC-MS.

1.2.4 Determining P450 concentration using CO binding

Concentration of active P450 in cell lysate was determined in 96 well format. Lysate (100 µL) was mixed with 100 µL Kpi (1 M, pH 8.0), containing 0.08 M sodium dithionite. Difference in absorption was measured before and after CO binding at 490 nm and 450 nm.

1.3 Protein expression and purification

1.3.1 Protein expression in 96 well plates

Expression of $P450_{BM3}$ variants in 96 well plates and preparation of cell lysates were performed as described.[1]

1.3.2 Protein expression in shake flasks (100 mL-3 L)

E.coli DH5α was used for expression of P450 monoxygenase variants which were cloned into pcWORI. *E.coli* BL21 was used for expression of P450 heme domains which were cloned into pET22b+ providing a C-terminal His6-tag. LB medium was inoculated from a single colony and grown overnight at 37 °C in a shaker. Two liters of TB medium supplemented with 100 µg/mL of ampicillin was then inoculated 1/100 with the preculture, and incubated in a 37 °C shaker until the OD600 of the solution reached 1.6 -1.8 (about 5 hours). At this point, the culture was cooled to 25 ˚C; IPTG and δ-aminolevulinic acid (ALA) at final concentrations of 0.5 mM each were combined. The expression culture was then incubated at 25 °C in a shaker for 24 hours, centrifuged, and the pellet was frozen at -20˚C.

1.3.3 Protein purification

1.3.3.1 Protein purification for crystallization

The pellets from the 2 L culture containing the His-tagged heme domain were resuspended in a total volume of 80 mL His-Buffer A (20 mM Tris, 100 mM NaCl, 20 mM imidazole, pH 8.0). The resuspended cell solutions were then sonicated to lyse, and centrifuged at 25,000 rpm for an hour. The supernatant (containing the His-tagged protein) was then loaded in 25-40 mL batches onto a 5 mL Ni-NTA column (GE Healthcare) using the ÄKTA-FPLC system. The program consists of an equilibration step with 4 column volumes (CVs) of His-Buffer A, the injection of the sample solution, followed by a wash step with His-Buffer A for 2 CVs, a 6 CV step at 95% His-Buffer A and 5% His-Buffer B (20 mM Tris, 100 mM NaCl, 300 mM imidazole, pH 8.0), a 6 CV step at 46% His-Buffer A and 54% His-Buffer B, a 2 CV step at 100% His-Buffer B, and a 3 CV re-equilibration. The fractions containing the desired protein were then combined and concentrated using a 30 kDa Millipore filtration device at 3,000 rpm. These devices were centrifuged and refilled with Q-Buffer A (25 mM Tris, pH 8.0) three times to complete the buffer exchange. This protein solution was then loaded onto an ion-exchange column (Q-column) using the ÄKTA system again. The program used with Q-Buffer A and Q-Buffer B (25 mM Tris, 1 M KCl or NaCl, pH 8.0) consists of an equilibration with 4 CVs of Q-Buffer A, an injection of the sample solution, a 3 CV wash step with Q-Buffer A, a 6 CV step with 92% Q-Buffer A and 8% Q-Buffer B, followed by a 10 CV gradient to 50% Q-Buffer A and 50% Q-Buffer B, a 2 CV step with Q-Buffer B, and a 3 CV re-equilibration. The fractions containing most of the desired protein were concentrated to 2 mL and loaded onto a gel-filtration column in two increments, letting the ÄKTA run overnight at 0.5 mL/min to separate the protein from impurities of different size. The eluted protein was concentrated to 300-500 µM (15-25 mg/mL) and flash-frozen in 30-100 µL aliquots. 5 µL samples of protein were taken after each step and run on a SDS-PAGE protein gel to follow the purification.

1.3.3.2 Protein purification for bioconversions

Each of the pellets from a 500 mL culture containing the monooxygenase was resuspended in 20 mL Q-Buffer A, sonicated to lyse, and centrifuged at 25,000 rpm for an hour. The supernatant containing the monooxygenase protein was then loaded onto ion-exchange Q-column using the ÄKTA-FPLC system. The program consisted of a 4 CV equilibration step, a 2 CV wash step with Q-Buffer A, a 6 CV step with 12% Q-Buffer B, a 6 CV step with 34% Q-Buffer B, a 2 CV step with 100% Q-Buffer B, and a 2 CV re-

equilibration step. The fractions containing the desired protein were then combined and concentrated using a 30 kDa Millipore filtration device. These devices were centrifuged at 3000 rpm and refilled with 0.1 M Kpi buffer, pH 8.0 three times to complete the buffer exchange. The protein was then flash-frozen in aliquots and stored at -80 °C.

1.4 T_{50} measurements

The P450 variant (final concentration: 1.0 µM, volume: 180 µL) was incubated at the indicated temperatures for 15 min in Kpi buffer(pH 8.0, 100 mM) in a thermoblock, then placed on ice. Samples were centrifuged for 3 min at 14,000 rpm. The supernatant (160 µL) was carefully removed and used to determine the concentration of active P450 using CO-binding in a 96 well format. All measurements were done in duplicate. Results were plotted and the T_{50} value was determined from a logistic fit using the program SigmaPlot.

1.5 Crystallization

1.5.1 Conditions for crystallization

Following purification, the proteins were concentrated to 15-30 mg/mL and set up in seven different crystallization screens using the sitting drop method in a 96 well format using 150-300 nL per well. Only variants 13C9 and 22A3 yielded crystals in the conditions tried. The screens of 13C9 yielded crystals in some of the wells containing either polyethylene glycol (PEG) 3350 or PEG 8000 and either Tris-HCl at pH 8.5 or Bis-Tris at pH 6.5. The best conditions contained Li_2SO_4 or $MgCl_2$ as well as PEG 3350 and Tris-HCl. The screens of 22A3 yielded crystals in some of the wells containing sulfate and citrate ions, as long as PEG 3350 was present. HEPES buffer at pH 7.5 was present in some of the crystal-producing wells. The best condition for 22A3 contained sodium citrate and PEG 3350. Rescreens were set up using varying concentrations of the best conditions for each protein, increasing the yield of crystals. Crystals which appeared prismatic with well-defined edges were looped out of their wells to be transferred to an oil solution before being flash-frozen and shipped to the x-ray diffraction facility.

Numerous crystals were successfully harvested from the screens, and three of the 22A3 crystals (in sodium citrate and PEG 3350) diffracted, two with a resolution of 3.2–3.6 Å, and one with a resolution of 2.67Å. This dataset, however, was not complete.

1.5.2 Measuring crystals

X-Ray diffraction data were collected on beamline 12-2 at Stanford, CA using a MAR CCD detector.

1.5.3 Analysis of crystal structure

Data reduction was done with CCP4. The structure was solved by molecular replacement (PAHSER) as implemented in the CCP4 suite using the structure of1JPZ as a search template. The structure was further refined using REFMAC and to a final resolution of 3.1. Data collection and refinement statistics are shown in Suppl. Table 5.

1.5.4 Protein Data Bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank (www.pdb.org) with accession number **XXX**.

1.6 Modeling substrates

To model naproxen, ibuprofen, and naphthalene acetic acid, we began with OpenBabel[2] (version 2.2.3 http://openbabel.sourceforge.net/ Aug 17 2009) to derive 3-D coordinates, followed by Multiconf-DOCK to generate favorable conformers[3]. We used the resulting confomers to generate an ensemble of potential transition state conformations. Specifically, we used Python scripts to place each potential reaction site into the heme pocket according to the transition state geometry proposed by Ryde and coworkers[4]. Our template protein model was the structure of intermediate 22A3 (chain B). To create a transition state ensemble, we expanded upon each substrate conformer by varying the docking orientation (10° increments for up to 3 degrees of freedom). We discarded the majority of the resulting substrate poses that clashed with rigid portions of the enzyme (i.e. an atom was within 2.2 Angstroms of the backbone, heme atoms, a beta carbon, or a Proline sidechain atom). Finally, we used the SHARPEN molecular modeling library[5] and the all-atom Rosetta energy function[6] to further filter the remaining substrate poses. Specifically, we selected Rosetta atom types for the substrates and performed combinatorial rotamer optimization of amino acid sidechains neighboring the substrate. Prospective transition state conformations of the lowest energy were visually inspected in PyMOL[7]

1.6.1 Structural comparison of 22A3 with wild-type $P450_{BM3}$

We were interested to see if the 18 mutations had caused significant alterations in the 22A3 protein backbone compared to wild-type $P450_{BM3}$ and other mutants crystallized so far. Simple structural alignment of two arbitrarily chosen $P450_{BM3}$ structures however does not address this question adequately because P450s can show significant structural flexibility as a consequence of crystal packing and the presence of substrate. Comparisons between substrate-bound and substrate-free structures of $P450_{BM3}$ have revealed major conformational differences and given the first detailed picture of substrate-induced conformational fit in $P450_{BM3}$.[8] Therefore, it is difficult to determine which changes in structure can be attributed to mutations rather than structural flexibility.

To minimize the effect of regions known for structural flexibility, we sought to compare 22A3 to the closest related P450BM3 structure which should be in a similar conformation. Therefore, we compared 22A3 (chain B) to all available crystallographic models for $P450_{BM3}$ wild-type and variants. We used the program TMalign[9] to obtain the root mean square deviation of the backbone atoms after superposition. Our 22A3 model was most similar to 1FAG[8] (Suppl. Figure 2), a crystal structure of wild-type $P450_{BM3}$ with a fatty acid substrate, which was used for further comparison. We then tried to identify which regions of 22A3 diverge from the existing structure 1FAG. Specifically, we used the program ProFIT (Martin, A.C.R. and Porter, C.T., http://www.bioinf.org.uk/software/profit/) to report the deviations for each residue after global structure alignment. Suppl. Figure 2 shows the deviations of 22A3 (chain B) backbone atoms from 1FAG (chain A) for each amino acid residue.

1.7 Synthetic procedures

1.7.1 Reagents and analytical methods.

Chemical reagents, substrates and solvents were purchased from Sigma-Aldrich, Acros Organics, Fluka, and Kaironkem. Silica gel chromatography purifications were carried out using AMD Silica Gel 60 230-400 mesh. HPLC analyses were carried out using a Waters 2690 instrument equipped with a 996 Photodiode Array Detector or a Beckman System Gold with Solvent Module 126 and Detector Module 168. For each HPLC system, either a Phenomenex Gemini 5u C 18 column or a Phenomenex Luna 5u C18 column was used. HPLC-MS analyses were carried out using an Agilent 1100 Series LC/MSD device and Kromasil 100, 5 um C 18 column. ^{1}H and ^{13}C NMR spectra were recorded on a Varian Mercury 300 spectrometer (300 MHz and 75 MHz respectively). 2D-NMR spectra were recorded on an INOVA-600 spectrometer. NMR spectra are internally referenced to residual solvent peak. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s=singlet, d=doublet, t=triplet, q=quartet, m=multiplet, br=broad), coupling constant (Hz), integration). Data for ^{13}C are reported in the terms of chemical shift (δ ppm). High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility.

1.7.2 Bioconversions

Bioconversions typically contained 1–3 µM $P450_{BM3}$ variant, 1 mM substrate, 1 % DMSO in Kpi buffer (100 mM, pH 8.0). Reactions were started by addition of a cofactor regeneration system containing final concentrations of 0.5 mM NADPH, 20 mM glucose-6-phosphate, and 1 U/mL glucose-6-phosphate dehydrogenase. Bioconversion were shaken (up to 1 mL scale) or stirred (6 to 300 mL scale) at 4 °C for 16 hours.

1.7.2.1 Conversion of 1 to 2

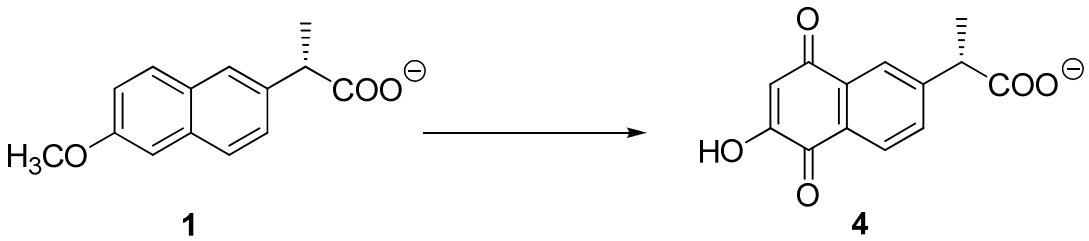
1 **2**

Conversion of 1 to 2. 138 mg (0.6 mmol) **1** were dissolved in 6 mL DMSO and then added to 300 mL of Kpi buffer (100 mM, pH 8.0). To this mixture, a final concentration of 6 µM P450 variant X3H1 and a cofactor regeneration system (final conc.: 500 µM NADPH, 30 mM glucose-6-phosphate, 1 U/mL glucose-6-phosphate dehydrogenase) were added and stirred at 4 °C. After for 12 h, brine was added, and the pH was adjusted to 2.0. The product was extracted with methylene chloride (4 x 100 mL). The combined phases were dried over anhydrous Na_2SO_4 and evaporated under reduced pressure. The residue was purified using a preparation-scale HPLC Agilent 1200 instrument equipped with a Phenomenex Gemini 5u C 18 250x21.2 mm column (gradient 0-50% acetonitrile with 0.1% trifluoroacetic acid in H_2O) to yield **2** (55.6 mg, 42.9 %, peach solid).

^{1}H-NMR ($CDCl_3$, 300 MHz): δ 1.52 (d, 3H, CH_3), δ 3.82 (q, 1H, CH), δ 7.03 (d, 1H), δ 7.07 (s, 1H), δ 7.38 (d, 1H), δ 7.60 (d, 1H), δ 7.64 (s, 1H), δ 7.68 (d, 1H). ^{13}C-NMR (75 MHz, DMSO): δ 20.1, δ 47.8, δ 110.9, δ

120.7, δ 128.1, δ 128.2, δ 128.7, δ 131.1, δ 131.6, δ 138.0, δ 157.8, δ 179.8. EI^+: exact mass calculated for $C_{13}H_{12}O_3$ requires m/z 216.0786, found 216.0797.

1.7.2.2 Conversion of 1 to 4

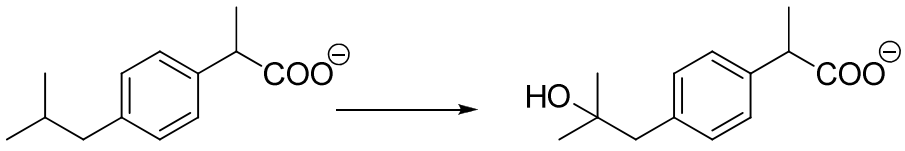


1 **4**

Conversion of 1 to 4. 55.2 mg (0.24 mmol) **1** were dissolved in 2.4 mL DMSO and then added to 240 mL of Kpi buffer (100 mM, pH 8.0). To this mixture, a final concentration of 3 µM P450 variant X3H1 and a cofactor regeneration system (final conc.: 500 µM NADPH, 30 mM glucose-6-phosphate, 1 U/mL glucose-6-phosphate dehydrogenase) were added and stirred at 4 °C. After for 12 h, brine was added, and the pH was adjusted to 2.0. The product was extracted with methylene chloride (4 x 100 mL).The combined phases were dried over anhydrous Na_2SO_4 and evaporated under reduced pressure. The residue was purified using flash chromatography (gradient 2-25% methanol in methylene chloride) to yield **4** (11.9 mg, 20 %, red-orange solid).

2D-NMR (CD_3OD): gCOSY, gHSQCAD, gHMBC, ^{1}H-NMR (600 MHz) δ 1.54 (d, 3H, CH3), δ 3.34 (q, 1H, CH), δ 5.29, δ 7.37, δ 7.60, δ 7.72. ^{13}C-NMR (150 MHz) δ 18.7, δ 48.7, δ 105.5, δ 123.0, δ 123.7, δ 128.5, δ 129.4, δ 135.7, δ 151.7, δ 172.4, δ 176.2, δ 180.9, δ 187.4. ^{13}C-NMR (DMSO, 150 MHz): δ 19.58, δ 49.50, δ 106.26, δ 123.99, δ 124.57, δ 129.08, δ 129.33, δ 135.58, δ 151.92, δ 172.38, δ 176.20, δ 180.66, δ 187.29. TOF MS ES^-: exact mass calculated for $C_{13}H_9O_5$ requires m/z 245.0450, found 245.0462.

1.7.2.3 Conversion of 5 to 6



5 **6**

Conversion of 5 to 6. 71.6 mg (0.37 mmol) **5** were dissolved in 3.6 mL DMSO and then added to 180 mL of Kpi buffer (100 mM, pH 8.0). To this mixture, a final concentration of 4 µM P450 variant W7D8 and a cofactor regeneration system (final conc.: 500 µM NADPH, 30 mM glucose-6-phosphate, 1 U/mL glucose-6-phosphate dehydrogenase) were added and stirred at 4 °C. After for 12 h, brine was added, and the pH was adjusted to 2.0. The product was extracted with methylene chloride (4 x 100 mL).The combined phases were dried over anhydrous Na_2SO_4 and evaporated under reduced pressure. The residue was purified using flash chromatography (gradient 0-20% ethyl acetate with 0.2% formic acid in methylene chloride) to yield **6** (74.3 mg, 96 %, white powder).

^{1}H-NMR ($CDCl_3$, 300 MHz): δ 1.22 (s, 6H, CH_3), δ 1.51 (d, 3H, CH3), δ 3.72 (q, 1H, CH), δ 7.18 (d, 2H), δ 7.27 (d, 2H). ^{13}C-NMR ($CDCl_3$, 75 MHz): δ 18.12, δ 29.15, δ 44.86, δ 70.87, δ 127.41, δ 130.76, δ 136.82, δ 137.96, δ 179.88. HRMS (EI+): exact mass calculated for $C_{13}H_{18}O_3$ requires m/z 222.1256, found 222.1279.

1.7.3 Sample preparation for HPLC and HPLC-MS

For use with HPLC or HPLC-MS, 100 µL samples of the bioconversions were taken and mixed with 100 µL acetonitrile. The samples were then centrifuged and filtered into HPLC vials. In HPLC analysis, chromatographic separation was achieved using a Waters 2690 Separations Module fitted with a reverse-phase C8 Luna column (150 x 4.60 mm) with 5 µm particle size. Buffer A was 0.2 % formic acid in water, buffer B was acetonitrile (HPLC-grade). Separation was achieved by using a linear gradient starting from 90% A/10 % B and going to 10% A/90% B over a period of 16 min, with a flow rate of 0.6 mL/min. Detection was achieved using a Waters 996 Photodiode Array Detector. In HPLC-MS analysis, 10 µL glacial acetic acid was added to each sample prior to loading. Chromatographic separation was achieved using a Kromasil 100, 5 um reverse-phase C18 column. Detection and fragmentation were achieved using an Agilent 1100 Series LC/MSD device. Buffer A was 0.1% acetic acid in water, and buffer B was methanol, with a flow rate 0.2 mL/min.

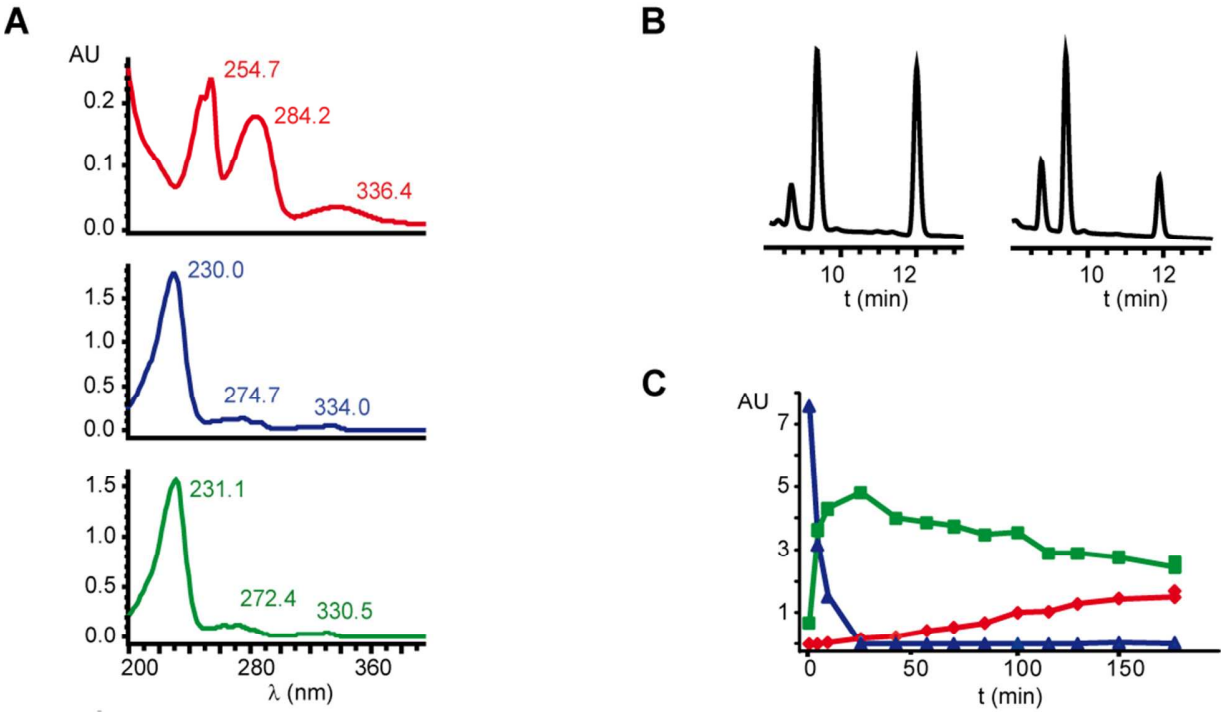
1.7.4 Product Isolation

To work up larger-scale (150-300 mL) bioconversions, at least 100 mL brine were added. The solution was then titrated to pH 2-3 to protonate the carboxylic acid and make it go to the organic phase. 50 mL methylene chloride was added, before filtering the solution through sand and celite. The organic phase was separated, and the aqueous phase was extracted with three quantities of 50 mL CH_2Cl_2 and one quantity of 10 mL ethyl acetate. The organic phases were combined and dried with Na_2SO_4. The solution was then gravity filtered into a 1 L round-bottom flask and concentrated under reduced pressure to 1-5 mL. To separate reactants from products, flash chromatography was used with a solvent gradient from 100 % methylene chloride to 90/10 % methylene chloride/methanol. Thin-layer chromatography was used to determine the components of the fractions, which were respectively combined and concentrated under reduced pressure to dryness. Purified samples were then analyzed using tandem mass spectroscopy (MSMS) in the Agilent 1100 Series LC/MSD device, and both ^{13}C- and ^{1}H-NMR in $CDCl_3$ in a Varian Mercury 300 MHz NMR instrument.

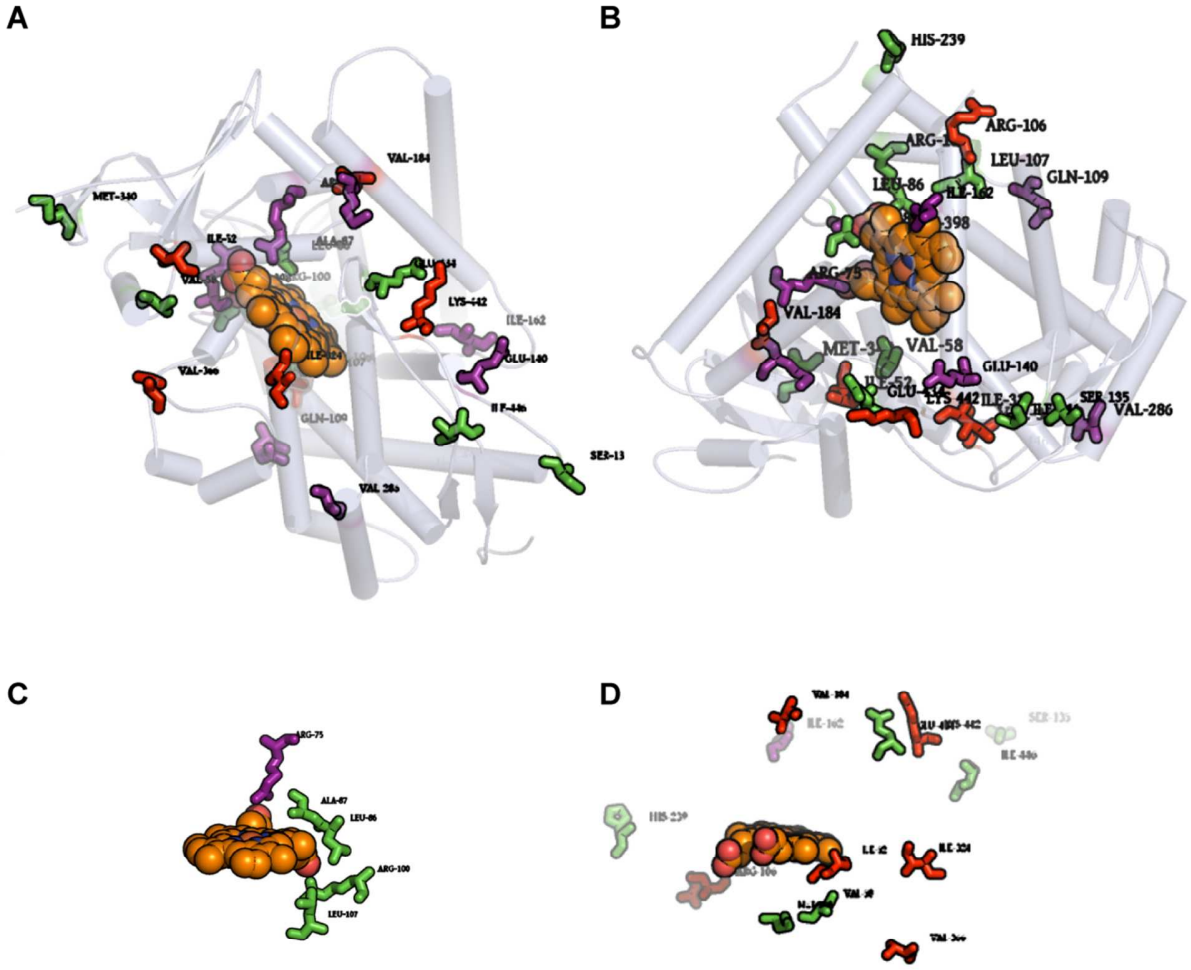
2 Supplementary Information

Supplementary Table 1 Sequences of the $P450_{BM3}$ variants

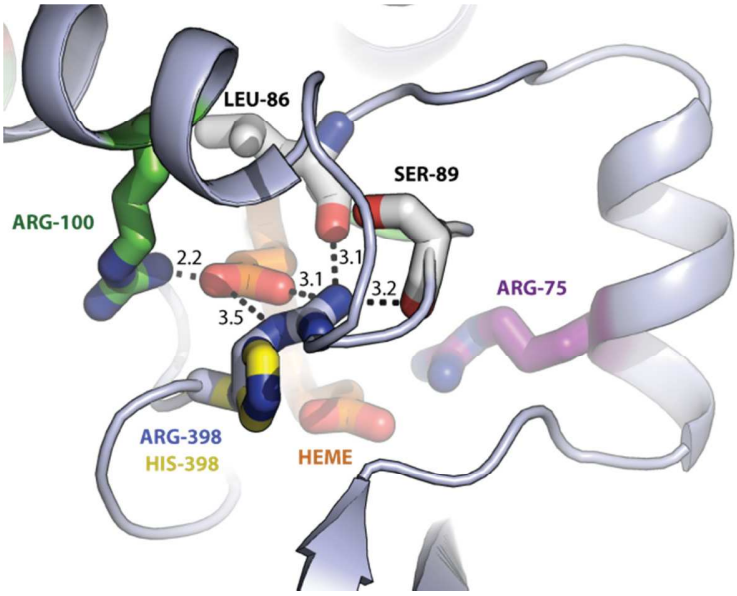
Enzyme	Class	Amino acid mutations compared to wildtyp $P450_{BM3}$
5H6	Peroxygenase	L52I, I58V, L86L, F87A, H100R, S106R, F107L, A135S, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
13C9	Peroxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, A135S, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
13C9R1	Monooxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, A135S, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
22A3	Monooxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, A135S, F162I, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
22A3his6	Peroxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, A135S, F162I, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
16G2	Monooxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, Q109L, A135S, E140G, F162I, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
W7D8	Monooxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, Q109L, A135S, E140G, F162I, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I
X3H1	Monooxygenase	L52I, I58V, L75R, L86L, F87A, H100R, S106R, F107L, Q109L, A135S, E140G, F162I, A184V, N239H, S274T, V286A, L324I, V340M, I366V, R398H, K434E, E442K, V446I



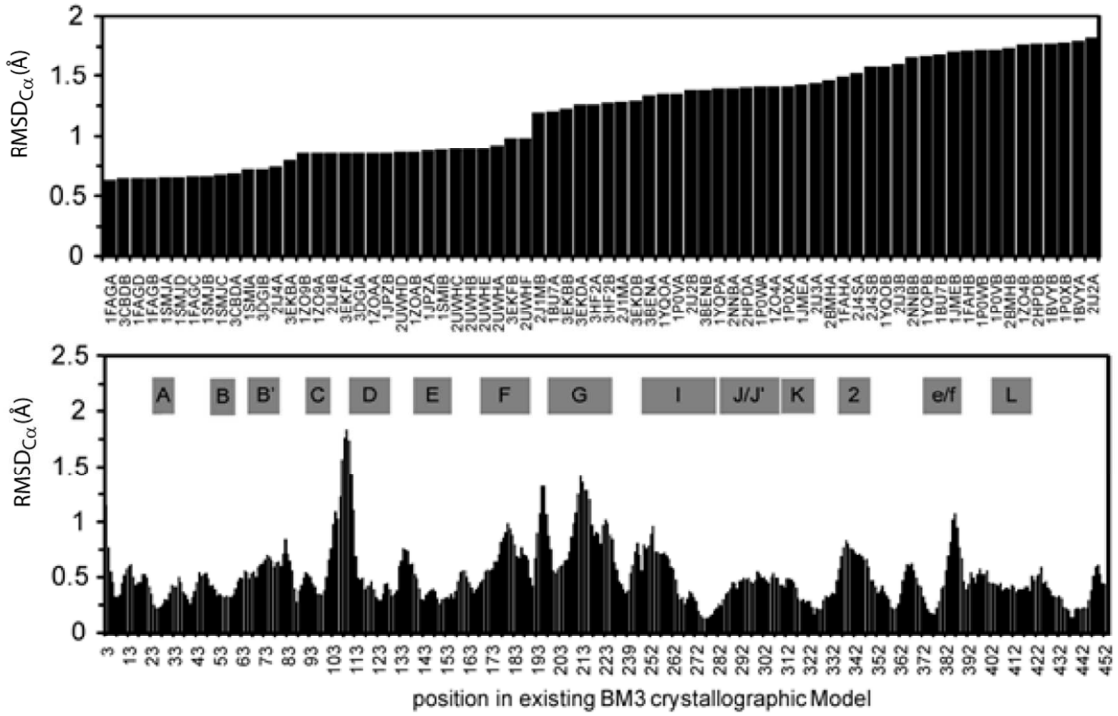
Supplementary Figure 1 A) Variants improved on naproxen lead to formation of an additional product with a different absorption spectrum. Top panel (red): adsorption spectrum of additional product, middle panel (blue): absorption spectrum of desmethylnaproxen, lower panel (green): absorption spectrum of naproxen. B) Conversions of naproxen (left) and desmethylnaproxen (right) with $P450_{BM3}$ variant X3H1 yield the same product. Thus, this additional product can be formed from desmethylnaproxen. C) Time course of naproxen conversion with $P450_{BM3}$ variant X3H1. Naproxen can be fully converted to desmethylnaproxen before the additional product becomes prominent. Formation of the additional product decreases the concentration of desmethylnaproxen, suggesting formation from desmethylnaproxen rather than naproxen.



Supplementary Figure 2 Structure of $P450_{BM3}$ variants highlighting mutations compared to WT. A) Mutations of X3H1 compared to WT. Structure of 22A3 highlighting mutations compared to WT and residues mutated in X3H1. B) Like A) but rotated 90/0/90°. C) D) Mutations of 22A3 compared to wild-type $P450_{BM3}$ C) Mutations within 8 Å of heme: L75R, F87A, H100R, F107L. D) Mutations outside 8 Å of heme: L52I, I58V, S106R, A135S, F162I, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I. Color coding: Red: thermostabilizing mutations[10], green: mutations from wt to 21B3[11] (except for M145V which was reverted during thermostabilization), magenta: mutations of 22A3 compared to 5H6 which introduce activity on naproxen, ibuprofen, 1-naphthalene acetic acid.



Supplementary Figure 3 Crystal structure of 22A3 highlighting the charge network environment of Arg 75, Arg 100, and Arg 398. Mutation of the highly conserved arginine 398 to histidine in W7D8 is likely to influence the conformation of the heme due to loss of the bidentate coordination of the heme propionate – although His 398 is likely to retain a hydrogen bond to the propionate as shown in the modeled histidine rotamer (yellow). The R398H mutation is also likely to disrupt the conformation of the loop following the B' helix. Specifically, hydrogen bonds to the backbone carbonyls of Leu 86 and Ser 89 will be lost. By incurring conformational change of the heme or the loop, the R398H mutation could indirectly interact with Arg 75.



Supplementary Figure 4 Upper panel: Structural similarity of 22A3 to known $P450_{BM3}$ variants: RMSD of αCs of 22A3 from existing $P450_{BM3}$ crystallographic models reveals chain A of 1FAG as most closely related structure. Lower Panel: Residual deviations of 22A3 αCs from 1FAG Chain A.

2.1 Alternative starting points for directed evolution

Other $P450_{BM3}$ variants from our freezer collection (a total of 800 active and expressed $P450_{BM3}$ mutants based on wild-type and 9-10A[12]) were screened on naproxen to find a potentially better starting point for directed evolution. The best mutant was 9-10A having 5 additional mutations: A78V, A82W, V184A, L188W, A328F which gave 0.5 % conversion in a reaction containing 0.5 µM naproxen and 3 µM purified enzyme, corresponding to a TTN of ~1.

2.2 Libraries generated for directed evolution

Supplementary Table 2 Libraries generated for directed evolution on naproxen.

Round	Parent	Fraction of library with <10% CO binding signal	Average mutation rate (nucleotides per heme domain)
1	13C9R1	48%	3.3
2	22A3	65%	4.7
3	16G2	65%	6.1
4	16G2	38%	1.5
5	W7D8	20%	2.3

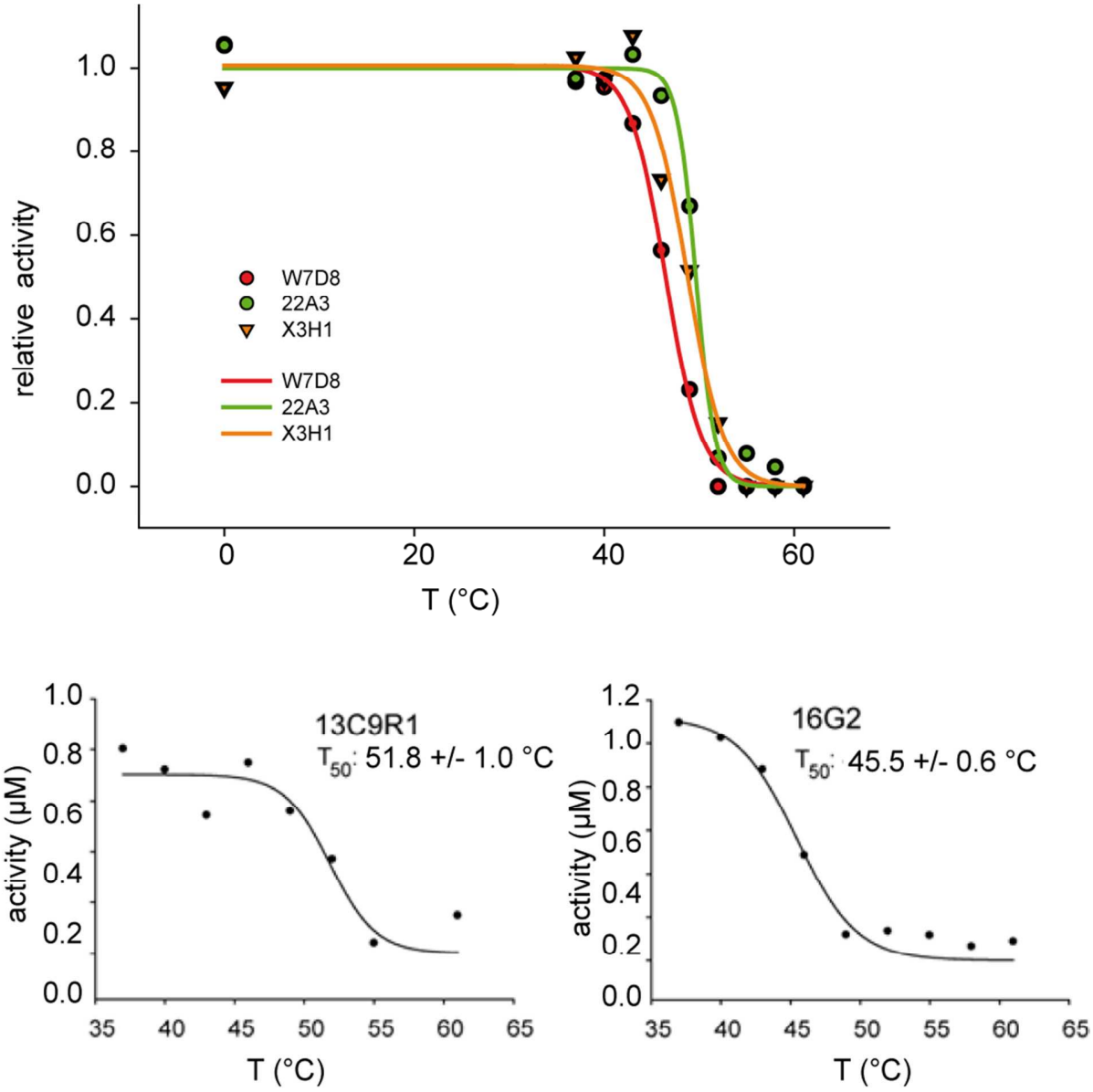
2.3 Improved variants from directed evolution on naproxen

Supplementary Table 3 Results of directed evolution on naproxen. Winners after rescreen chosen for purification and further evaluation in bioconversions. All listed winners showed higher turnovers than their parent. Winners chosen as parent for next round are typed in bold.

Round	parent	Winners after rescreen	Mutations compared to parent
1	13C9R1	22E7	E143G
		22A3	**F162I**
		11B5	D182V, S305R, H409Q
		26B6	E64V, E294D
		6E3	V413I
		6F10	T49A, P170S, M185K
2	22A3	**16G2**	**Q109L, E140G**
		RS3E5	D231G
		RS3A3	I122V, P170T
		RS2E10	Q110H, D233E
3	16G2	16G2	
4	16G2	**W7D8**	**V286A, R398H**[1]
5	W7D8	**X3H1**	**M185K**

[1] W7D8 also contains 3 additional silent nucleotide mutations: T351C, A423G, A795G.

2.4 Thermostabilities



Supplementary Figure 5 Thermostability measurements and T_{50} evaluation of $P450_{BM3}$ mono-oxygenase variants.

2.5 Importance of mutations

Supplementary Table 4 Mutations of 22A3 and their putative effect based on comparison of crystal structures for 22A3 and wild-type

Mutation	**Putative effect**
L52I (stabilizing)	moving to β-branched amino acid is favorable in β-sheet
I58V	
F87A	
H100R	(R398H might be coupled to H100R because these residues are interacting with the same heme-propionate)
S106R (stabilizing)	structural deviations
A135S	increasing hydrophilicity in exposed group; no interactions with other residues in monomer
A184V (stabilizing)	in helix F; affects packing with helix B' (these 2 helices move closer together)
N239H	Surface
S274T	in helix I; additional methyl in hole; P441 is closest amino acid and moves a little bit further away
L324I (stabilizing)	in helix K, neighbours conserved and not moving, conservative
V340M (stabilizing)	increasing hydrophobic packing
I366V (stabilizing)	removal of methyl
K434 and E 442K (stabilizing)	charge reversal on long hairpin. 2nd mutation (E442K) completes polarity change and is therefore thermostabilizing. Crystal structures show that 434 and 442 interact directly
V446I	no neighbours moving, added bulk, should be stabilizing
Later mutations:	
Q109L	still mostly exposed but more buried. Mutation to L helps interaction with non-polar protein interior. Could be coupled to mutations at 106/107 where biggest structural deviation (C/D helix)
R398H	perturbation of charge network that compensates the heme. R398H could reduce high temperature factor in B' helix. Communicating with R75 through heme and through the loop that follows helix B'. R398H might be coupled to H100R because these residues are interacting with the same heme-propionate
M185K	on surface, conservative, potential for direct interaction with naproxen even in the transition state, depending on exact placement of helix F.

2.6 Crystal structure

Supplementary Table 5 Data collection and refinement statistics

Data collection	
P450 construct	$P450_{BM3}$ variant 22A3
No. of crystals	1
Complex	none
Space group	P222
Unit cell (a, b, c) Å	82.35, 83.77, 143.02
α, β, γ	90.0, 90.0, 90.0
SSRL beamline	12-2
Wavelength (Å)	1.000
Resolution (Å)	3.1 (3.27)
Total observation	115370
Unique reflections	17710
Completeness %	96.1 (94.0)
$\langle I/\sigma_I \rangle$	10.8 (3.0)
R_{sym} (*I*)	13.9 (60.6)
Refinement	
R_{cryst}	21.6 (33.6)
R_{free} (5% of data)	28.0 (35.2)
Root mean square deviation bonds (Å)	0.01
Root mean square deviation angles (°)	1.41
Average B-factor (Å^2)	83.2

Refinement Statistics were determined using phenix.validate

[1] M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442-13450.
[2] R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, E. L. Willighagen, *J. Chem. Inf. Model.* **2006**, *46*, 991-998.
[3] N. Sauton, D. Lagorce, B. O. Villoutreix, M. A. Miteva, *BMC Bioinformatics* **2008**, *9*, 184.
[4] P. Rydberg, L. Olsen, P. Norrby, U. Ryde, *J. Chem. Theory Comput.* **2007**.
[5] a) I. V. Loksha, J. R. Maiolo, C. W. Hong, A. Ng, C. D. Snow, *J. Comput. Chem.* **2009**, *30*, 999-1005; b) K. M. S. Misura, D. Chivian, C. A. Rohl, D. E. Kim, D. Baker, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5361-5366.
[6] C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, in *Methods in Enzymology, Vol. 383* (Eds.: B. Ludwig, L. J. Michael), Academic Press, **2004**, pp. 66-93.
[7] L. Schrodinger, Version 1.3 ed.
[8] H. Y. Li, T. L. Poulos, *Nat. Struct. Mol. Biol.* **1997**, *4*, 140-146.
[9] Y. Zhang, J. Skolnick, *Nucleic Acids Res.* **2005**, *33*, 2302-2309.
[10] O. Salazar, P. C. Cirino, F. H. Arnold, *ChemBioChem* **2003**, *4*, 891-893.
[11] P. C. Cirino, F. H. Arnold, *Angew. Chem., Int. Ed.* **2003**, *42*, 3299-3301.
[12] P. Meinhold, M. W. Peters, M. M. Y. Chen, K. Takahashi, F. H. Arnold, *ChemBioChem* **2005**, *6*, 1765-1768.